Abstract Title Page

Title:

Geographic Variation of District-Level Gender Achievement Gaps within the United States

Authors and Affiliations:

Sean Reardon, Stanford University
Erin Fahle, Stanford University
Demetra Kalogrides, Stanford University
Anne Podolsky, Learning Policy Institute
Rosalia Zarate, Stanford University

Abstract Body

Background / Context:

Gender achievement gaps on national and state assessments have been a popular research topic over the last few decades. Many prior studies examine these gaps in different subjects (e.g., mathematics, reading, and science) and grades (typically kindergarten through eighth grade) for students living in various regions (typically states or countries) using a variety of measures. Despite drawing different conclusions about the exact magnitude of the gaps, the general findings about the existence of gender achievement gaps from prior research are consistent. In mathematics, the majority of studies do not find a significant achievement gap when children enter kindergarten (Fryer & Levitt, 2009; Husain & Millimet, 2009; Lee, Moon, & Hegar, 2011; Robinson & Lubienski, 2011; Sohn, 2012), but show the existence of an average male-favoring gap in mathematics by the end of kindergarten lasting through fifth grade (Fryer & Levitt, 2009; Lee, Moon, & Hegar, 2011; Robinson & Lubienski, 2011; Penner & Paret, 2008; Sohn, 2012). Moreover, researchers have shown that male students are systematically overrepresented among the high achievers, yielding a meaningful "upper tail" gap in mathematics that may have important consequences for students' future engagement in the field (Robinson & Lubienski, 2011; Penner & Paret, 2008). In reading or English Language Arts (ELA), most studies find significant achievement gaps favoring girls beginning in kindergarten and persisting throughout elementary school of approximately 0.15 SDs, which are even larger among the lowest performing students (Chatterji, 2006; Fryer & Levitt, 2009; Husain & Millimet, 2009; Robinson & Lubienski, 2011). Overall, this research shows that estimated achievement gaps in the United States tend to follow a pattern in line with common stereotypes – that boys are better at mathematics and girls are better at reading.

Only a limited amount of this research, however, has focused on characterizing or explaining the variation of gender achievement gaps in mathematics and reading within the United States. Hyde et al. (2008) provide evidence that gaps in gender performance on state mathematics assessments vary across states for second through eleventh grades. Pope and Sydnor (2010) investigate the variation across states in the ratios of male-to-female students scoring above the 95th percentile in mathematics and science and the equivalent female-male ratio in reading. In their analysis, the authors find considerable variation in these high-end ratios, and further that regions with high-end gaps favoring males in mathematics and science are accompanied by high-end gaps favoring females in reading. In addition, the authors find that the variation in gender achievement gaps across different regions of the United States is best explained by the variation in regional stereotypes and gender attitudes, rather than socioeconomic characteristics, such as a region's average household income or proportion of individuals' with high school degrees. In other words, the authors conclude that students tend to perform in a stereotypical pattern when they live in communities that hold traditional views about gender abilities and roles, illustrated in part through survey data about the proportion of individuals in a community who think that "women are better suited for home" and "math is for boys" (Pope & Sydnor, 2010, p. 105).

Significantly more research on the geographic variation of gender achievement gaps has been done in the international context, looking across countries. These studies suggest that local social systems and population characteristics may influence the size of national gender achievement gaps (Hyde & Mertz, 2009; Guiso et al., 2008; Penner, 2008; Andreescu et al., 2003), yielding a promising avenue forward for research within the U.S.

Purpose / Objective / Research Question / Focus of Study:

In this paper, we seek to better understand the variation of gender achievement gaps within the United States by focusing at a higher geographic resolution than prior studies. We analyze gender achievement gaps within school districts across the nation. And, given the evidence of the state-level variation and the support for suspecting that gender achievement gaps vary amongst districts within a state, we investigate how local social and economic factors, local demographics, and school district characteristics may relate to the magnitude and direction of gender achievement gaps, with a specific focus on the extent to which the gaps appear in a "stereotypical" or "gender-favoring" pattern.

Setting:

This study focuses on student achievement in the United States. The data used in this study come from the EdFacts state accountability tracking system, which was provided to our team via restricted license by the National Center for Education Statistics (NCES).

Population / Participants / Subjects:

The population of study is U.S. school districts serving third through eighth grade students. The EdFacts database includes data for all (approximately 13,700) public school districts in the US, and contains math and ELA test score data in grades 3-8 from 2009-2012. In total, the data include roughly 200 million test scores. We are able to estimate gender achievement gaps in approximately 9,400 districts (we exclude very small—mostly rural—districts where gap estimates are extremely imprecise). These districts collectively enroll roughly 98% of public school students in the U.S. The population of districts, therefore, is representative of all but the smallest districts in the U.S.

Intervention / Program / Practice:

This study is a secondary data analysis of the EdFacts data with the intent of characterizing how gender achievement gaps vary in magnitude and direction across the U.S.

Research Design:

This paper analyzes gender achievement gaps within school districts across the nation and the local factors that vary with the gaps. Given the evidence of variation in the size (and direction, in the case of mathematics) of achievement gaps among school districts, we conduct a descriptive analysis to investigate whether gender gaps are associated with socio-demographic characteristics of the population. Our reasoning here is that gender gaps may, in part, be a product of local norms and expectations that are shaped both by socio-demographic characteristics and gender roles.

We adopt a two-dimensional framework for characterizing patterns in reading and mathematics achievement. The first dimension characterizes the gaps' alignment with traditional gender stereotypes—that boys outperform girls in mathematics and girls outperform boys in reading. The second dimension characterizes the extent to which both gaps, instead, tend to favor

one gender over the other. These two dimensions of gender achievement differences can be visualized using a coordinate plane, where a district's reading gap is plotted on the x-axis and its mathematics gap on the y-axis (please insert figure 1 here).

Our motivation for adopting this framework—or more specifically, jointly analyzing the mathematics and ELA gender gaps—is twofold. Based on prior research, we believe that there are two types of factors affecting the gender achievement gap: (1) factors that may lead to a general favoring of boys or girls in both subjects; and, (2) factors that may lead to more-or-less stereotypical achievement.

Data Collection and Analysis:

The achievement data used in this study come from the EdFacts state accountability tracking system described above. The EdFacts data are provided as counts of students scoring at the state proficiency levels (e.g. "Below Basic," "Basic," etc.) disaggregated by district, gender, grade, subject, and year. The data are available for the 2008-09 through 2011-12 school years, for the third through eighth grades in mathematics and ELA in the United States. We include all schools that serve elementary and middle school students, regardless of whether they are part of an elementary (K-8) or unified (K-12) district. Note that charter schools are included in this analysis, either as part of the public school district to which they are chartered, or we assign them to the nearest public district based on physical location.

To estimate the district-level achievement gaps from this coarsened data, we use the V-statistic (Ho, 2009; Ho & Haertel, 2006; Ho & Reardon, 2012). We estimate math and ELA gaps in each grade-by-year cell (in each of 6 grades in each of 4 years) within each district, and then average these 24 gaps within each district and subject. Roughly 85% of the variation in gaps is between districts, so this averaging discards little information.

We consider four different categories of covariates as correlates of district-level male-female achievement gaps: 1) socioeconomic differences between males and females residing in the district; 2) average socioeconomic characteristics of all residents in the district; 3) school district racial and economic composition; and 4) district-level male-female differences in behavioral outcomes. We construct measures of socioeconomic differences and average socioeconomic characteristics from the 2006-2010 5-year American Community Survey data. Our measures, estimated from the Common Core of Data (CCD), of school racial and economic composition include the percent black, Hispanic and white students in public schools as well as the percent eligible for free lunches. Finally, we use the 2010-2011 Civil Rights Data Collection (CRDC), which includes data from every public school in the nation, which include detailed information about student grade retention and suspension rates.

We first estimate the within-state bivariate correlations between the achievement gap measures and the covariates. In estimating the correlations, we take into account the fact that the achievement gaps are measured with error.[†]

[†] To estimate these within-state correlations we first center the estimated gaps and the covariates around their unweighted state means. We then estimate, via maximum likelihood, the true variance τ of the centered gaps. We then compute the correlation of the gaps with each covariate, weighting each observation by $(\hat{\tau} + \hat{v}_{ds})^{-1}$, where $\hat{v}_{ds}$ is the estimated sampling variance of the gap estimate $\hat{G}_d^s$.

We then estimate bivariate and multivariate associations between gender gaps and school district covariates using a set of models of the form:

$$\hat{G}_d^s = \alpha + \mathbf{X}_d\mathbf{B} + \mathbf{\Gamma}_{state} + u_{ds} + \epsilon_{ds}; \quad u_{ds} \sim N[0, \tau], \epsilon_{ds} \sim N[0, v_{ds}]$$

where $\hat{G}_d^s$ is an estimated pooled gender gap measure in district d and subject s; $\mathbf{X}_d$ is a vector of district covariates; $\mathbf{\Gamma}_{state}$ is a vector of state fixed effects; u_{ds} is an independent normally-distributed mean-zero residual with variance τ to be estimated; and ϵ_{ds} is an independent normally-distributed mean-zero sampling error term with known variance v_{ds} equal to the sampling variance of $\hat{G}_d^s$. This model describes the gap in each district as a function unobserved state characteristics (including differences in tests across states), a vector of district covariates, a residual error term indicating the difference between the true gap and that predicted by the covariates, and a second error term due to sampling error in the estimated gap. It is essentially a weighted fixed effects model, where each observation is given a weight equal to $(\tau + v_{ds})^{-1}$.

Findings / Results:

Among the roughly 9,400 districts in our sample, the average gender achievement gap in mathematics is near 0, although there is considerable variation across districts. In contrast, the average gap in ELA is roughly 0.25 SDs, in favor of girls. Again there is significant variation in the district-level gaps; however, they nearly always favor females. Plotting these gaps, we find that there is considerable variation among districts along both the gender-favoring and stereotype dimensions. (please insert figure 2).

The four different categories of covariates of district-level male-female achievement gaps explain most of the variation along the stereotype dimension, but little of the variation along the gender-favoring dimension. Moreover, we find that communities with more stereotypical gender gaps tend to be more white, higher income, and more educated.

Conclusions:

This district-level study provides critical insight into the variation of gender achievement gaps in the U.S., as well as an understanding of how correlates that are proximal to the students, rather than the more distal climate measures at the state or regional level used in prior literature, are associated with different patterns, specifically stereotypical or gender-favoring patterns, in the gaps.

Appendices

Appendix A. References

Andreescu, T., Gallian, J. A., Kane, J. M., & Mertz, J. E. (2008). Cross-Cultural Analysis of Students with Exceptional Talent in Mathematical Problem Solving. *Notices of the AMS*, *55*(10), 1248–1260.

Beller, M., & Gafni, N. (2000). Can Item Format (multiple choice vs. open-ended) Account for Gender Differences in Mathematics Achievement?. *Sex Roles, 42*(1-2), 1-21.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender Differences in Multiple-Choice Tests: the Role of Differential Guessing Tendencies. *Journal of Educational Measurement*, *28*(1), 23-35.

Chatterji, M. (2006). Reading Achievement Gaps, Correlates, and Moderators of Early Reading Achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, *98*(3), 489.

DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education, 11*(3), 279-299.

Fryer Jr, R. G., & Levitt, S. D. (2009). *An empirical analysis of the gender gap in mathematics* (No. w15430). National Bureau of Economic Research.

Gamer, M., & Engelhard Jr, G. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science 320*(5880), 1164.

Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, *34*(2), 201-228.

Ho, A. D., & Haertel, E. H. (2006). Metric-free measures of test score trends and gaps with policy relevant examples (CSE Report No. 665). Los Angeles, CA: *Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies.*

Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal "proficiency" categories. *Journal of Educational and Behavioral Statistics*, *37*(4), 489-517.

Husain, M., & Millimet, D. L. (2009). The mythical "boy crisis"? *Economics of Education Review*, (28), 38–48. doi:10.1016

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, *106*(22), 8801–8807.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494-495.

Lee, J., Moon, S., & Hegar, R. L. (2011). Mathematics skills in early childhood: Exploring gender and ethnic patterns. *Child Indicators Research*, *4*(3), 353-368.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, *136*(6), 1123.

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, (37), 239–253.

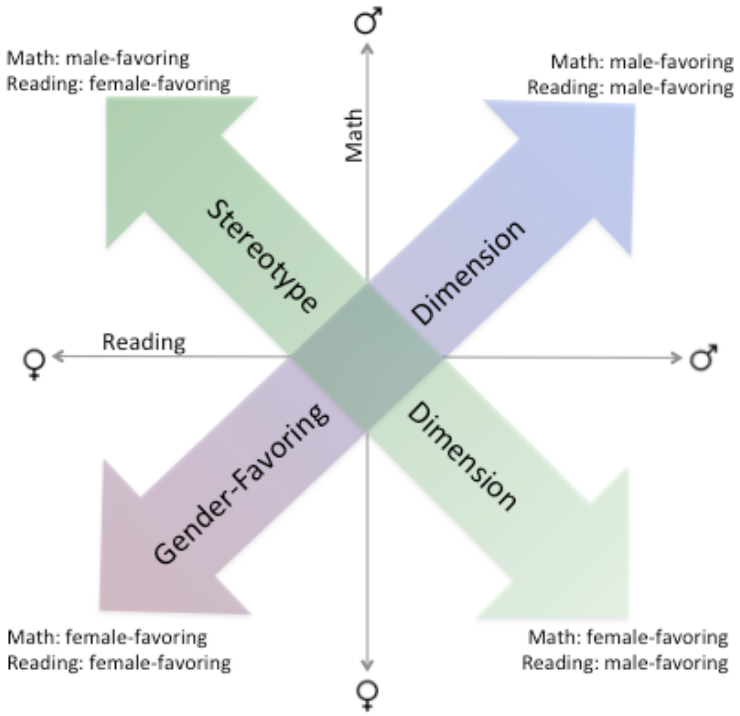
Pope, D. G., & Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *The Journal of Economic Perspectives*, *24*(2), 95-108.

Robinson, J. P., & Lubienski, S. T. (2011). The Development of Gender Achievement gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings. *American Educational Research Journal*, *48*(2), 268–302.

Sohn, K. (2012). A new insight into the gender gap in math. *Bulletin of Economic Research, 64*(1), 135-155.

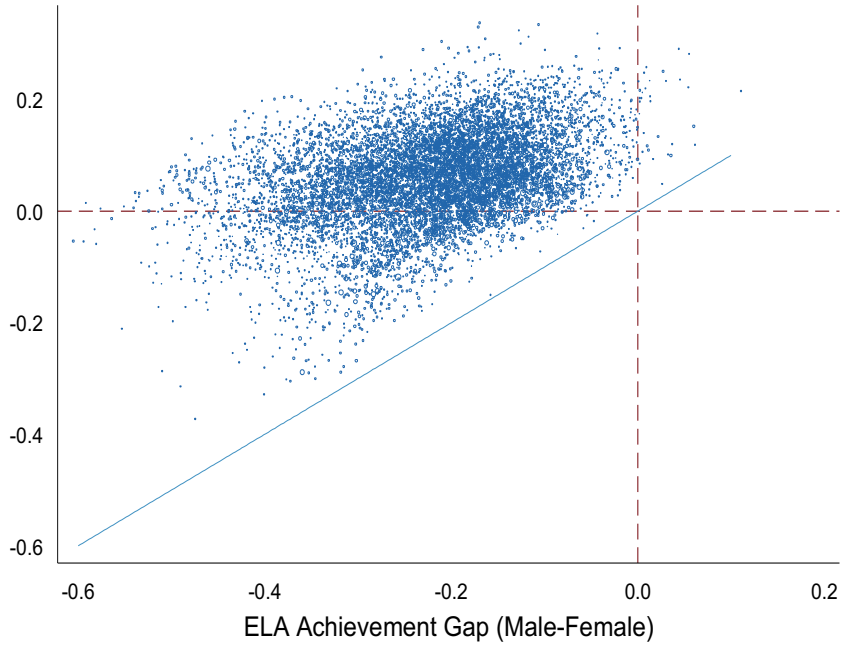
Appendix B. Tables and Figures

Figure 1: Dimensions of Gender Achievement Gaps



Notes: Reading gaps are plotted on the x-axis and corresponding mathematics gaps on the y-axis. Positive (negative) values indicate gaps are male-favoring (female-favoring). Gender equality in the gaps is at the origin.

Figure 2: Male-Female Mathematics and ELA Achievement Gaps, School Districts 2009-2012



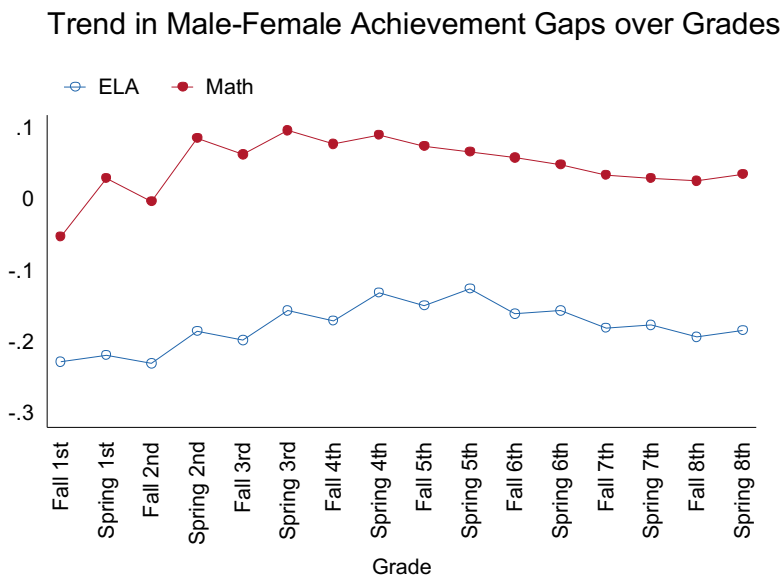
Notes: Reading gaps are plotted on the x-axis and corresponding mathematics gaps on the y-axis. Positive (negative) values indicate gaps are male-favoring (female-favoring). The model used to estimate average gaps includes state fixed effects and adds the average state NAEP gap to the Empirical Bayes estimate.

Table 1: Relationship between Proportion Multiple-Choice Items on State Tests and the Size of Gender Gaps, State-Level

	Mathematics				ELA			
	Grade 4		Grade 8		Grade 4		Grade 8	
Model 1: State-Level NAEP Audit Test								
Proportion Short Response+	-0.135	**	-0.109		-0.223	*	-0.376	**
Extended Response	(0.041)		(0.075)		(0.084)		(0.113)	
Model 2: District-Level NWEA Audit Test								
Proportion Short Response+	-0.126		-0.151	*	-0.296	**	-0.389	***
Extended Response	(0.122)		(0.068)		(0.098)		(0.101)	

All models are weighted by $1/se^2$. Standard errors that are clustered by state. Model 1 includes data from 2009 Ed Facts and NAEP data from grades 4 and 8. Model 2 data from 2009 Ed Facts and NWEA data sources from grades 4 and 8. The models are restricted to state or district by grade cells with gap data from both Ed Facts and NAEP/NWEA. Both models also include the proportion of "other" (not shown) items.

Figure 3: Trend in Male-Female Achievement Gaps over Grades



Notes: Estimated trends are taken from 3-level precision weighted HLM models with a non-parametric grade-by-term